

# yli130\_Assignment1

Yanxi Li

9/12/2021

## Read a file

```
# import the data
data <- read.csv('Movie-Ratings.csv')
```

## Summary Statistics

```
# overall the data
summary(data)
```

Film	Genre	Rotten.Tomatoes.Ratings..
Length:562	Length:562	Min. : 0.0
Class :character	Class :character	1st Qu.:25.0
Mode :character	Mode :character	Median :46.0
		Mean :47.4
		3rd Qu.:70.0
		Max. :97.0
Audience.Ratings..	Budget..million...	Year.of.release
Min. : 0.00	Min. : 0.0	Min. :2007
1st Qu.:47.00	1st Qu.: 20.0	1st Qu.:2008
Median :58.00	Median : 35.0	Median :2009
Mean :58.83	Mean : 50.1	Mean :2009
3rd Qu.:72.00	3rd Qu.: 65.0	3rd Qu.:2010
Max. :96.00	Max. :300.0	Max. :2011

```
str(data)
```

```
'data.frame': 562 obs. of 6 variables:
 $ Film      : chr  "(500) Days of Summer " "10,000 B.C." "12 Rounds " "127 Hours" ...
 $ Genre     : chr  "Comedy" "Adventure" "Action" "Adventure" ...
 $ Rotten.Tomatoes.Ratings..: int  87 9 30 93 55 39 40 50 43 93 ...
 $ Audience.Ratings..      : int  81 44 52 84 70 63 71 57 48 93 ...
 $ Budget..million...      : int   8 105 20 18 20 200 30 32 28 8 ...
 $ Year.of.release         : int  2009 2008 2009 2010 2009 2009 2008 2007 2011 2011 ...
```

```
# rename the column name
colnames(data) <- c('Film', 'Genre', 'CriticRating', 'AudienceRating', 'BudgetMillions', 'Year')
# check the column name
colnames(data)
```

```
[1] "Film"          "Genre"          "CriticRating"   "AudienceRating"
[5] "BudgetMillions" "Year"
```

```
# change 3 columns into factor
data$Film <- as.factor(data$Film)
data$Genre <- as.factor(data$Genre)
data$Year <- as.factor(data$Year)

# check the summary again
summary(data)
```

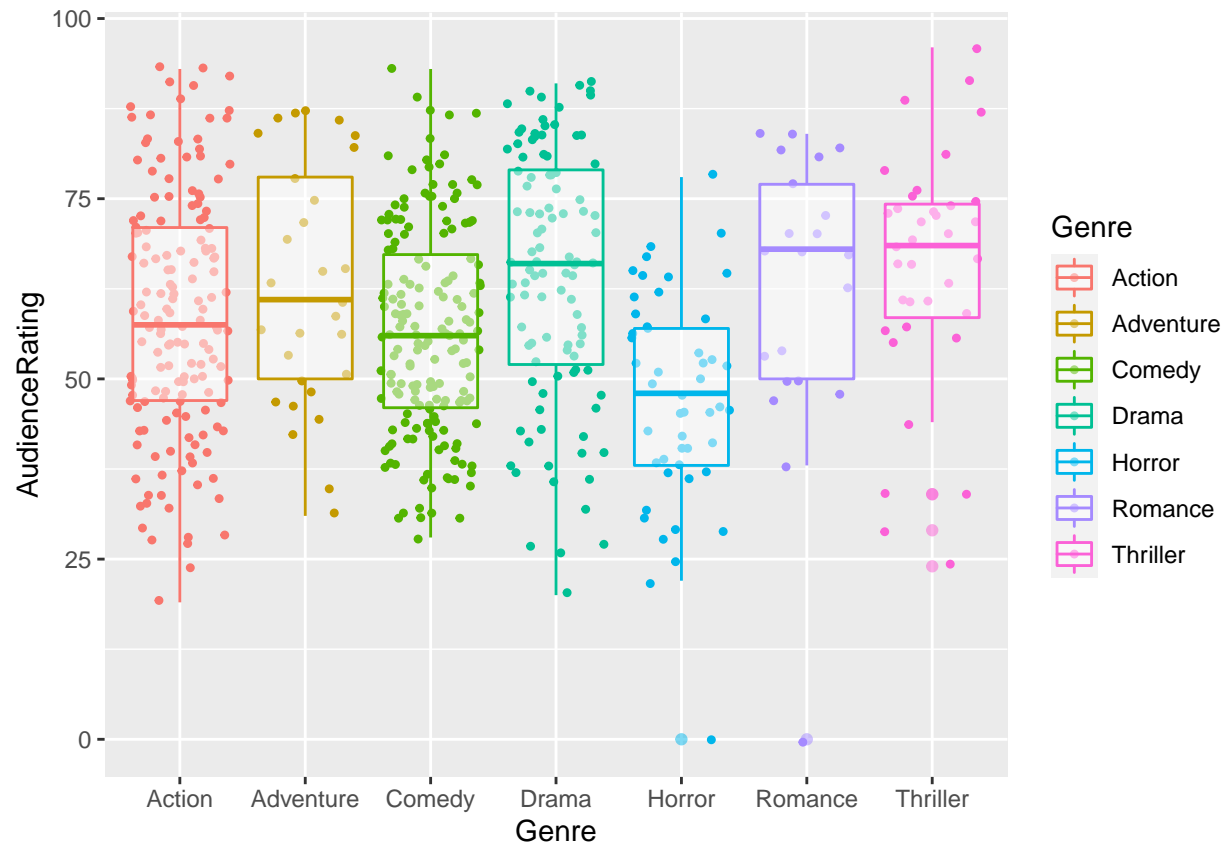
```

      Film      Genre      CriticRating      AudienceRating
(500) Days of Summer : 1   Action   :154   Min.    : 0.0   Min.    : 0.00
10,000 B.C.          : 1   Adventure: 29   1st Qu.:25.0   1st Qu.:47.00
12 Rounds            : 1   Comedy   :172   Median  :46.0   Median  :58.00
127 Hours             : 1   Drama    :101   Mean    :47.4   Mean    :58.83
17 Again              : 1   Horror   : 49   3rd Qu.:70.0   3rd Qu.:72.00
2012                  : 1   Romance  : 21   Max.    :97.0   Max.    :96.00
(Other)              :556   Thriller : 36
BudgetMillions      Year
Min.    : 0.0   2007: 79
1st Qu.: 20.0   2008:125
Median  : 35.0   2009:116
Mean    : 50.1   2010:119
3rd Qu.: 65.0   2011:123
Max.    :300.0
```

plot graphs

```
library(ggplot2)  # to plot the figure

#Audience_Rating & Genre
p <- ggplot(data=data, aes(x=Genre, y=AudienceRating, color=Genre))
p + geom_jitter(size=0.9) + geom_boxplot(alpha=0.5)
```



*#Critic\_Rating & Genre*

```
q <- ggplot(data=data, aes(x=CriticRating, y=AudienceRating, color=Genre))
q + geom_point(size=0.8) + geom_smooth(fill=NA)
```

